AMENDMENT TO THE CLAIMS

1. (Currently Amendment)   A method of segmenting words into component parts, the method comprising:

a processor determining a mutual information score for a pair of graphoneme units, comprising a first graphoneme unit and a second graphoneme unit, using the probability of the first graphoneme unit appearing immediately after the second graphoneme unit, the unigram probability of the first graphoneme unit and the unigram probability of the second graphoneme unit, each graphoneme unit comprising at least one letter in the spelling of a word;

a processor using the mutual information score to combine the first and second graphoneme units into a larger graphoneme unit; and

in a dictionary comprising segmentations of words into sequences of graphoneme units, a processor replacing the first and second graphoneme units with the larger graphoneme unit in each sequence of graphoneme units in which the first graphoneme unit appears immediately after the second graphoneme unit.

2. (Previously Presented)   The method of claim 1 wherein combining graphoneme units comprises combining the letters of each graphoneme unit to produce a sequence of letters for the larger graphoneme unit and combining phones of each graphoneme unit to produce a sequence of phones for the larger graphoneme unit.

3. (Original)   The method of claim 1 further comprising using the segmented words to generate a model.

4. (Original)   The method of claim 3 wherein the model describes the probability of a graphoneme unit given a context within a word.

5. (Original)   The method of claim 4 further comprising using the model to determine a pronunciation of a word given the spelling of the word.

6. (Previously Presented)   The method of claim 1 wherein using the mutual information score comprises summing at least two mutual information scores determined for a single larger graphoneme unit to form a strength.

7. (Previously Presented)   A computer-readable storage medium having computer-executable instructions stored thereon that when executed by a processor cause the processor to perform steps comprising:

    determining mutual information scores for pairs of graphoneme units found in a set of words, each graphoneme unit comprising at least one letter and each mutual information score for a pair of graphoneme units based on the probability of one graphoneme unit of the pair of graphoneme units appearing immediately after the other graphoneme unit of the pair of graphoneme units, and the unigram probabilities of each graphoneme unit in the pair of graphoneme units;

    combining the graphoneme units of one pair of graphonome units to form a new graphoneme unit based on the mutual information scores; and

    updating a segmentation of a word comprising a set of graphoneme units for the word that includes the pair of graphoneme units by replacing the pair of graphoneme units in the segmentation with the new graphoneme unit.

8. (Previously Presented)   The computer-readable storage medium of claim 7 wherein combining the graphoneme units comprises combining the letters of the graphoneme units to form a sequence of letters for the new graphoneme unit.

9. (Previously Presented) The computer-readable storage medium of claim 8 wherein combining the graphoneme units further comprises combining the phones of the graphoneme units to form a sequence of phones for the new graphoneme unit.

10. (Previously Presented) The computer-readable storage medium of claim 7 further comprising identifying a set of graphonemes for each word in a dictionary.

11. (Previously Presented) The computer-readable storage medium of claim 10 further comprising using the sets of graphonemes identified for the words in the dictionary to train a model.

12. (Previously Presented) The computer-readable storage medium of claim 11 wherein the model describes the probability of a graphoneme unit appearing in a word.

13. (Previously Presented) The computer-readable storage medium of claim 12 wherein the probability is based on at least one other graphoneme unit in the word.

14. (Previously Presented) The computer-readable storage medium of claim 11 further comprising using the model to determine a pronunciation for a word given the spelling of the word.

15. (Previously Presented) The computer-readable storage medium of claim 7 wherein combining graphoneme units based on the mutual information score comprises summing at least two mutual information scores associated with a new graphoneme unit.

16. (Currently Amended) A method of segmenting a word into syllables, the method comprising:

a processor segmenting a set of words into phonetic syllables using mutual information scores wherein using a mutual information score comprises computing a mutual information score for two phones by dividing the probability of two phones appearing next to each other in the set of words by the unigram probabilities of each of the two phones appearing in the set of words;

a processor using the segmented set of words to train a syllable n-gram model; and

a processor using the syllable n-gram model to segment a phonetic representation of a word into syllables via forced alignment.

17. (Currently Amended)   A method of segmenting a word into morphemes, the method comprising:

a processor segmenting a set of words into morphemes using mutual information scores wherein using mutual information scores comprises computing a mutual information score for two letters based on the probability of the two letters appearing next to each other in the set of words and the unigram probabilities of each of the two letters appearing in the set of words;

a processor using the segmented set of words to train a morpheme n-gram model; and

a processor using the morpheme n-gram model to segment a word into morphemes via forced alignment.